MOOK NATTAKAN

Data scientist

+6143-415-5648
nattakan.o1998@gmail.com
Blackburn South, Melbourne
Portfolio
LinkedIn

PROFESSIONAL SUMMARY

Data Scientist with a Master of Data Science (Distinction) from Deakin University and a Bachelor of Computer Engineering from Mahidol University. Experienced in end-to-end data projects spanning machine learning, data engineering, and deployed analytics dashboards using Python, SQL, GCP, Terraform, and Kestra, with a background in full-stack web development. Holds full working rights in Australia (Subclass 485 Graduate visa).

EMPLOYMENT HISTORY

Data Scienctist
Mar 2025 – Oct 2025

Chameleon, Melbourne

- Led an 11-person team delivering a geospatial parking study and a sleep pattern analysis on 10,115 participants
- Designed a structured data validation process and delivered a stakeholder poster presentation

Mobile App Developer
Jan 2022 – Oct 2022

Owat Pro and Quick Company Limited, Bangkok

- Led requirements gathering across 3 teams for a React Native workforce management app, selected over a commercial vendor
- Designed Figma wireframes and built the clock-in/out system before project handover

Web Developer Intern
May 2019 – Aug 2019

TeleInfo Media Public Company Limited, Bangkok

- Built an internal revenue and workforce reporting dashboard using Vue.js, MongoDB and Elasticsearch
- Delivered an Excel BI analysis projecting new customer acquisition targets per salesperson

SKILLS

Programming & Querying: Python, Advanced SQL, JavaScript

Data Science & Statistical Analysis: scikit-learn, Random Forest, SHAP, pandas, statistical modeling

Data Science Theory & Methodology: hypothesis testing, feature engineering, model evaluation & validation, bias-variance tradeoff, cross-validation

Artificial Intelligence & Deep Learning: LSTM, neural network design, CNNs

Data Engineering & Cloud: BigQuery, GCP, dbt, Terraform, Kestra, Bruin, AWS

Data Visualization & BI: Looker Studio, seaborn, Folium

Computer Vision: OpenCV, object detection, OCR, image preprocessing

Databases: MySQL, PostgreSQL

Web Development: Vue.js, Node.js, React Native, Figma

Tools & Version Control: Git, Vercel, Trello

PROJECTS

Retail Analytics Pipeline | dbt · BigQuery · GCP · Looker Studio

- Designed a 5-model dbt and BigQuery ELT pipeline processing 541K transactions (£8.18M revenue) into a star schema with staging, fact, dimension, and mart layers, end-to-end from raw CSV to dashboard-ready analytics in a single dbt run
- Delivered a Looker Studio dashboard revealing UK market dominance (£7.5M of £8.18M total revenue), year-end seasonality (top day £200K on Dec 9, 2011), and high customer concentration insights directly supporting retention and inventory planning decisions

Spotify Track Popularity Analysis | Python · XGBoost · scikit-learn · Pandas · seaborn · SHAP

- Benchmarked Linear Regression, Random Forest, and XGBoost across 113,843 tracks; Random Forest achieved $R^2 = 0.51$, RMSE = 0.70, outperforming XGBoost ($R^2 = 0.26$), confirming non-linear feature interactions drive popularity
- Used SHAP analysis to identify duration, speechiness, energy, and valence as the strongest predictors; found pop-film and k-pop as the highest-popularity genres, surfacing actionable label A&R insights

Melbourne Parking Accessibility Analysis | Python · pandas · Folium · geopy · Open Data API

- Integrated 5 City of Melbourne open datasets via REST API, applying a 400m geospatial proximity model to classify 9 healthcare and education landmarks into demand risk tiers; identified Carlton Primary and Kensington Primary Schools as high-risk parking shortage zones
- Found Kensington Primary had limited bus and tram access, directly informing a recommendation to increase public transport frequency as an alternative to new parking infrastructure; delivered findings via an interactive Folium risk map for urban planning stakeholders

Population-Level Sleep Pattern Analysis | Python · MNE · LSTM · pandas · NSRR SHHS dataset

- Built an LSTM model combining EEG sleep-stage sequences with demographic features from 10,115 participants; identified age as the strongest predictor of sleep efficiency ($r = -0.60$), with sex and BMI showing negligible effects
- Applied early stopping to prevent overfitting (optimal convergence at epoch 3–4) and produced a correlation heatmap across demographic features, translating clinical EEG data into actionable public health insights on age-related sleep deterioration

Facial Emotion Recognition | PyTorch · torchvision · ResNet · scikit-learn · TensorBoard

- Fine-tuned a pretrained ResNet CNN to classify facial expressions into 6 emotions (Happy, Sad, Fear, Anger, Disgust, Pain) across 1,198 images, achieving 0.797 ROC-AUC (macro) despite the small, imbalanced dataset
- Addressed class imbalance with weighted cross-entropy loss and data augmentation (random crop, horizontal flip); found Angry and Disgust were hardest to separate due to overlapping facial muscle patterns, and that adding model complexity worsened overfitting rather than improving results

Document Deskewing Pipeline | OpenCV · Tesseract OCR · Hough Transform

- Built a computer vision pipeline to detect and correct tilt in scanned/photographed documents, improving downstream OCR accuracy, took Tesseract output from unreadable noise to fully correct text in test cases
- Benchmarked 3 candidate-point selection strategies (all-pixel, centroid, baseline) across skew-detection thresholds, selecting a centroid-based approach that cut total pipeline runtime from ~0.14s to ~0.05s versus the naive method while maintaining accuracy

GPT-2 Fine-Tuning for Biomedical Text Generation | GPT-2 · LoRA · PyTorch · NLTK

- Fine-tuned GPT-2 on biomedical abstracts using full fine-tuning and LoRA, cutting perplexity from 98.28 to 17.11 while training 57x faster and updating only 0.24% of parameters with LoRA
- Found LoRA matched or beat full fine-tuning on both perplexity and BLEU while acting as a natural regularizer against overfitting on a small dataset

Web Application for Medical Distribution | Vue.js · Node.js · MySQL

- Built a full-stack web app to help manage and distribute medicines between hospitals and pharmacies
- Developed the frontend using Vue.js, focusing on ease of use and responsiveness across different screen sizes
- Built the backend services using Node.js and REST APIs to handle core functionality, with a MySQL database structure to store and manage all relevant data

EDUCATION HISTORY

• 2026 - Master of Data Science	Deakin University, Melbourne	Distinction Average
• 2021 - Bachelor of Computer Engineering	Mahidol University, Bangkok	GPA 3.34

CERTIFICATION

- HackerRank - Python, SQL
- DataTalks.Club - Data Engineering Zoomcamp